JUNE 1966
REPORT SM-48464-TPR-2

RESEARCH ON THE UTILIZATION OF PATTERN RECOGNITION TECHNIQUES TO IDENTIFY AND CLASSIFY OBJECTS IN VIDEO DATA

TECHNICAL PROGRESS REPORT NO.2
COVERING PERIOD FROM
1 OCTOBER 1965 THROUGH 20 MAY 1966

PREPARED UNDER CONTRACT NAS 12-30
FOR THE

ONAL AERONAUTICS & SPACE ADMINISTRATION
ELECTRONICS RESEARCH CENTER
CAMBRIDGE, MASSACHUSETTS

ASTROPOWER LABORATORY
2121 CAMPUS DRIVE • NEWPORT BEACH, CALIFORNIA

MISSILE & SPACE SYSTEMS DIVISION
DOUGLAS AIRCRAFT COMPANY, INC.
SANTA MONICA CALIF.

Report SM 48464-TPR-2

RESEARCH ON THE UTILIZATION OF PATTERN RECOGNITION TECHNIQUES TO IDENTIFY AND CLASSIFY OBJECTS IN VIDEO DATA

Technical Progress Report No. 2
Covering Period From
1 October 1965 through 20 May 1966

Prepared Under Contract NAS 12-30
for the
National Aeronautics and Space Administration
Electronics Research Center
Cambridge, Massachusetts

MISSILE & SPACE SYSTEMS DIVISION
ASTROPOWER LABORATORY
Douglas Aircraft Company, Inc.
Newport Beach, California

FOREWORD

This report was prepared by the Astropower Laboratory of the Douglas Aircraft Company, Inc., Newport Beach, California, in fulfillment of NASA contract NAS 12-30. The work was sponsored by the NASA Electronics Research Center, Cambridge, Massachusetts. Mr. Eugene Darling was the Technical Officer for this office.

The studies began on June 18, 1965. This document is the second periodic technical report.

Work on the contract was performed by the Electronics Laboratory of the Astropower Laboratory. The contributors were V. W. Goldsworthy, R. D. Joseph, C. C. Kesler, W. B. Martin, J. N. Medick, A. G. Ostensoe, R. G. Runge, and S. S. Viglione.

SUMMARY

This report documents the work accomplished during the period including the fourth through the eleventh months of the contract on the "Utilization of Pattern Recognition Techniques to Identify and Classify Objects in Video Data." The experimental results obtained on features of the lunar terrain are presented in detail.

Twelve design techniques, consisting of combinations of two methods for deriving property filter sets and six methods for deriving decision functions were utilized. Three recognition tasks were considered--separating craters with central elevations from craters without elevations, separating rima from wrinkle ridges, and separating the craters from the linear features. There are thus 36 system-task combinations.

One of the methods for generating property filter sets consistently provided higher performance levels than the other. In the assessment of the methods for designing decision functions, the results depended upon the task. On the most difficult task, for which performance levels were quite low, the two simplest methods gave the best results. On the task of intermediate difficulty, all six techniques were about equally effective. On the easiest task, two other systems were best, giving performance levels satisfactory for an operational system. Task difficulty appears to depend primarily on the size of the significant pattern features, and to some extent, on pattern definition.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1.0 INTRODUCTION

1.1 Purpose and Scope of the Program

The number of adaptive techniques for the design of pattern recognition methods has been proliferating. Comparisons of these techniques, however, are virtually nonexistent. In many cases, the experiments performed to validate individual techniques are difficult to evaluate due to their limited extent. Accordingly, the objective of this program is to assess the feasibility of adaptive design for pattern recognition for video data which might be acquired by a spacecraft.

A literature survey has been conducted, and was reported in the first periodic report.[1] Based on this survey, a number of adaptive design techniques were selected. These represent methods which appear to be most heavily reported, and which seem to bear the most promise.

Seven classes of patterns have been selected. Four of these are features of lunar topography and three are types of cloud cover occuring in NIMBUS records. The lunar features consist of

1. Craters with conspicuous central elevations
2. Craters without such elevations
3. Wrinkle ridges
4. Rima

This report is concerned with the experiments in which 12 combinations of adaptive design techniques are applied to three recognition tasks on the lunar features. The application of six techniques to two recognition tasks on the NIMBUS data will be reported in the final report.

1.2 Background Discussion

The adaptive design techniques are considered in the context of a design philosophy, which is illustrated in Figure 1. The process of recognition is composed of four parts – signal conditioning, known property extraction, statistical property extraction, and computation of the decision.

The purpose of signal conditioning or preprocessing is to build into the system the required invariances prior to the actual recognition, to

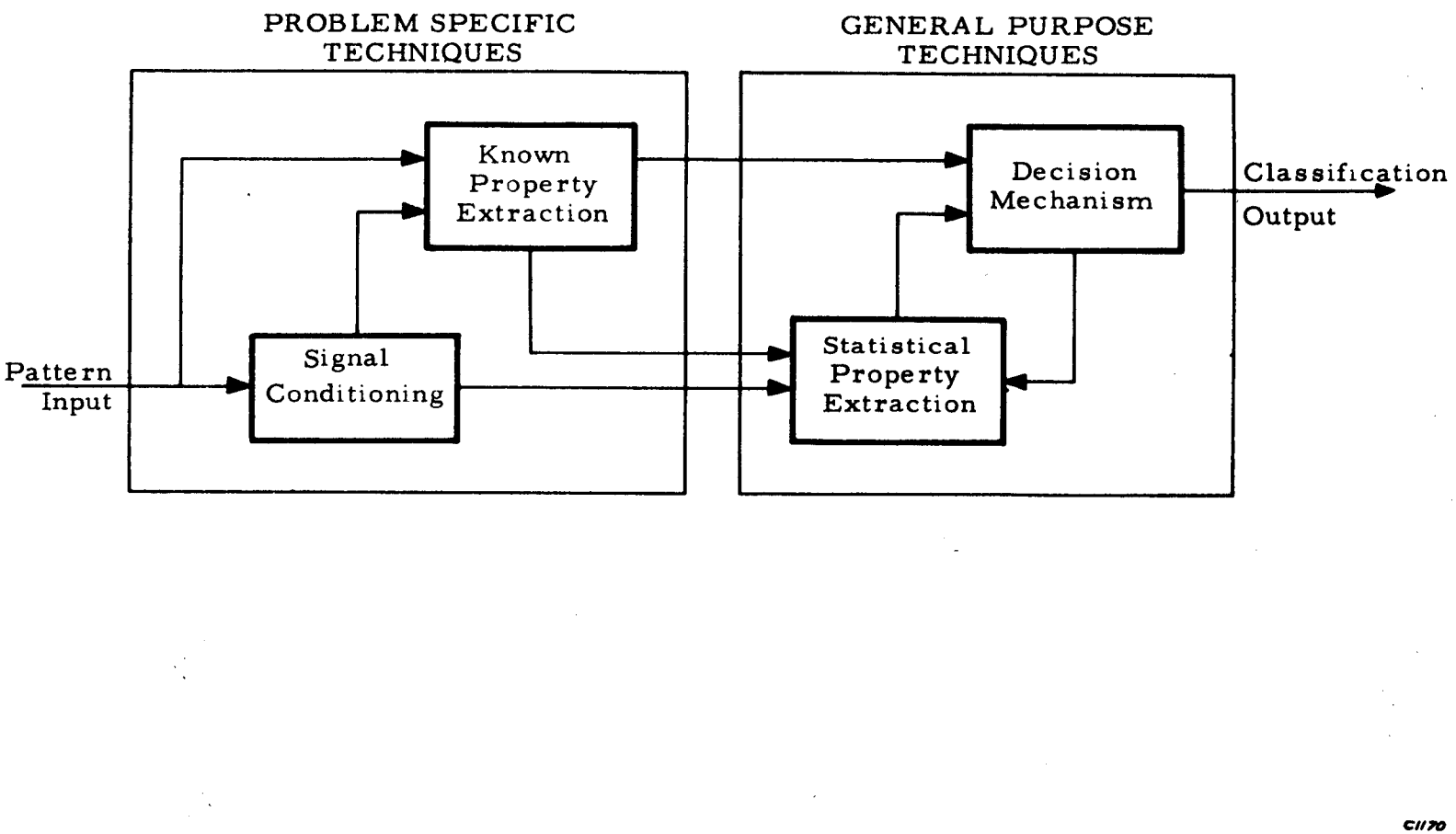


Figure 1. Pattern Recognition Mechanism

enhance the pattern to ease the recognition process, and to format the data for further processing. Techniques such as Laplacian filtering for edge enhancement, scanning, magnification control, autocorrelation, resolution regulation, and obtaining correlations against a variety of orthogonal and nonorthogonal functions are examples of signal conditioning methods which have been applied in pattern recognition.

From the conditioned signal, or from the raw data, properties which are known to be, or are suspected of being, of value in performing the recognition, are extracted. This phase of the design requires careful study and analysis of the patterns. Together with the signal conditioning phase, it provides the system designer an opportunity to transfer some of his knowledge and experience in performing the recognition tasks to the system. The results obtained are, however, specific to the task at hand and are not readily extended to new pattern recognition tasks.

The list of known properties may not be adequate for high performance on the task at hand, or may not be suitable for the nature of the decision function. In these cases, the list of properties must be expanded. This is accomplished by statistically analyzing samples of patterns of each class. The value of the properties derived from the analysis of these samples is dependent on the representativeness and size of the samples, and the signal to noise characteristics of the sample patterns.

The number of possible property profiles in general will be too large to permit cataloging them. Therefore, the classification to be assigned by the system to an unknown pattern must be computed from its property profile. It is in the area of specifying the decision function based on the sample patterns that most research in adaptive pattern recognition has been performed. For the most part, linear and piecewise linear decision functions have been considered. Research on the design of decision mechanisms and on statistical property extraction provides more general purpose results than does work done in signal conditioning and known property extraction.

For the experiments of this program, the signal conditioning consists of a limited degree of size and position normalization, resolution regulation, and contrast and brightness control. This work is described in Section 2

of this report. Known property extraction is not included in this effort, but is being investigated in a parallel study being performed in-house at NASA ERC. Two techniques for statistical property extraction are combined with six techniques for specifying a decision function. These are described in Section 3. The structure considered is shown in Figure 2. The experiments performed with the resulting 12 combinations on the lunar features are described in Section 4.

1.3 Summary of Work Performed

In this period, all of the required computer programs were written and debugged. All of the lunar and NIMBUS patterns were selected, recorded with a slow-scan TV, digitized, normalized, and edited. The experiments on the lunar data were completed. The hardware feasibility study was half completed. Only the work on the lunar data is reported here.

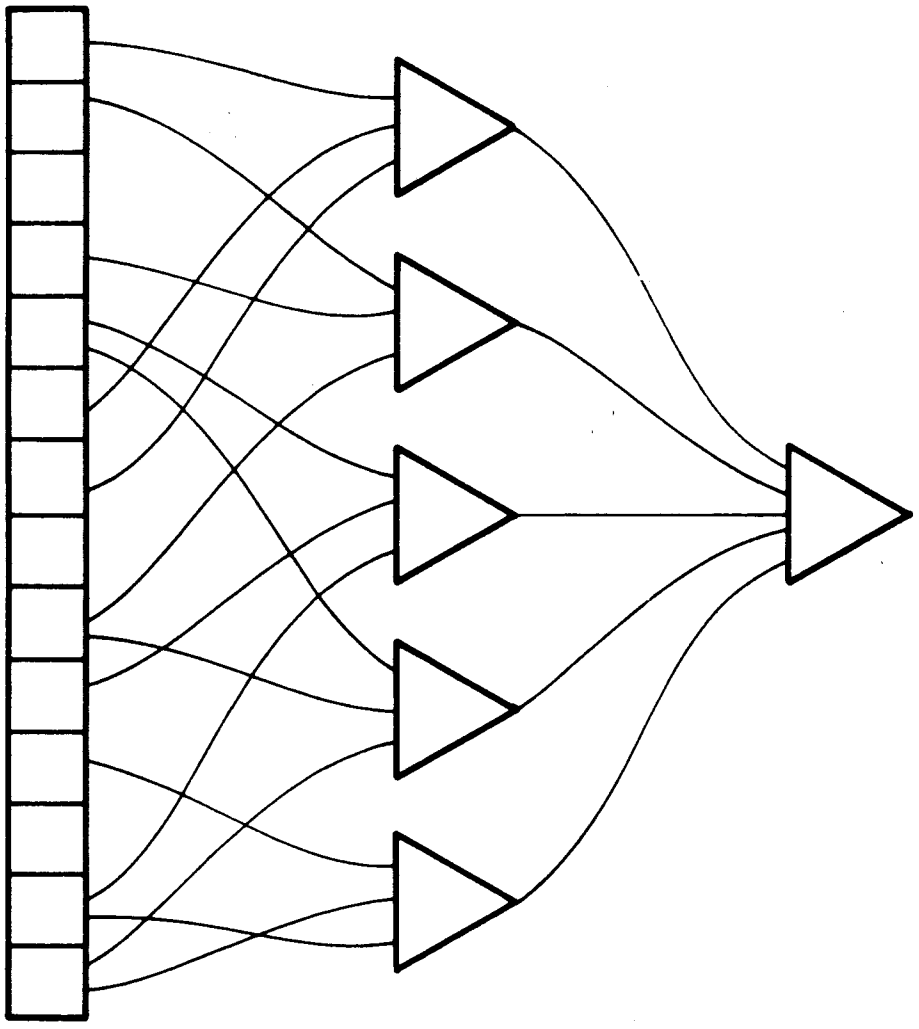


Figure 2. Decision Network Structure

2.0 GENERATION OF LUNAR PATTERN FILES

2.1 Origin of Pattern Sets

The initial pattern recognition experiments were based on a set of photographs of the lunar surface. The photographs were derived from Air Force lunar atlases,[2,3] originating at various astronomical observatories in the United States and France.

Mr. Eugene Darling, Technical Officer from NASA for this program, selected from these sources 198 lunar features categorized as follows:

Feature	No. of Cases
Craters, No conspicuous central elevation	53
Craters, One or more conspicuous central elevation	53
Rima (rilles or trenches)	52
Wrinkle ridges	40

These features formed the basic pattern set from which the training and independent generalization pattern samples were generated.

2.2 Preprocessing

The ultimate goal of this study is to develop a system which could be placed aboard a spacecraft to perform an examination of the optical patterns received and relay only selective information identifying the features or objects encountered. The amount of data that could be collected was vividly demonstrated by the outstanding performance of Surveyor I. A recognition system capable of this selective evaluation may provide a savings in weight, power and processing time for future interplanetary exploration vehicles. The recognition system would be required to respond correctly over a wide range of pattern size variations, pattern translations, and rotations and to compensate for variations in picture brightness. System designs to account for these variations could possibly be achieved by including in the training set of patterns all possible combinations of size, translation, and rotation of the features recorded at varying brightness levels. However this would require a very large pattern set and would significantly increase the time and expense needed to record and process these patterns.

As a result three assumptions were made to permit the use of a much smaller training sample in the simulation experiments without major adverse affects on the classification capability of the resulting recognition system. These resulted in a limited amount of pattern preprocessing consistent with the general design philosophy.

It was found, in examining the photographs, that in most cases human observers could not distinguish ridges from rima without knowledge of the direction of illumination from the sun. For example, if the sun is illuminating the moonscape from a relatively low angle, a ridge will be brightest on the side closest to the sun and cast a shadow on the side away from the sun. The opposite situation obtains for the rima or lunar canyon. While the craters could be recognized without prior knowledge of the sun's direction, the characteristic pattern of bright and dark crescents was tied directly to direction of illumination. Orienting the camera aboard the spacecraft to bear a constant relative angle with respect to the sun in the plane of the lunar surface was considered technically feasible since such craft presently use sun sensors to position and stabilize themselves. This allows the designer of the recognition system to take advantage of a priori knowledge of the direction of the sun and greatly reduces the number of possible pattern variations.

The second assumption concerned size normalization of the patterns. The craters chosen for this study varied in diameter from 10 KM to 100 KM. In order to classify the smallest crater in the set with an aperture adequately large to encompass the largest crater, many different examples of craters of the smallest size would have been required at various positions in the receptive field, to make the pattern statistically significant with respect to the rest of the picture. In addition, the results of a resolution investigation indicated that such an aperture would have required 250 x 250 points to provide a sufficient number of resolvable elements to classify the smallest craters in the set. To accommodate these pattern size variations a simple size normalization of the crater patterns was performed. The diameter of the largest crater was specified to be no greater than 1-1/2 times the diameter of the smallest crater in the set. Despite the restrictions imposed by this normalization it was felt that the applicability of the resulting machine design to the general recognition problem would not be limited. A recognition system with connections

specified by the experimental design could scan small sections of the received image to classify the smallest craters; the same design, with connections expanded by a fixed ratio about the center point of the scan aperture, could be used to scan larger areas and classify the larger craters. Alternately a zoom lens as used in Surveyor I could perform the size normalization. The linear features presented a more difficult problem in size normalization. The ridges and rima did not have a convenient dimension comparable to the diameter of the craters. They varied considerably both in width and length. As a compromise, it was decided to normalize the size of the linear features so that the distance across the feature was approximately 10% to 15% of the width of the aperture used for the craters. This achieved sufficient resolution across the feature to determine the light side and shadow side of the ridge or rima but retained its linear characteristic. In most cases this meant using only part of the length of the linear pattern.

The final assumption was that the amount of light entering the vidicon aboard a spacecraft could be controlled electronically to maintain a constant average brightness for the received pictures. (The use of a controllable iris on the cameras of Surveyor I clearly illustrates this technique.) This assumption permitted adjustment of the brightness levels of the digitized photographs to compensate for variations in video gain inherent in the slow scan television system used for conversion of the pattern photographs.

To obtain an estimate of resolution required for recognition of the lunar patterns, typical examples were selected from the four pattern groups and placed before the slow scan television camera. The vertical sweep was adjusted to give 160 lines/inch on the TV monitor. The monitor display was photographed as the distance between the TV camera and pattern was increased in five steps. The farthest distance was chosen to give about 15 lines of resolution across a crater. Examination of the monitor photographs indicated that human observers could accurately distinguish between craters without and craters with conspicuous elevations with about 25 lines of resolution across the crater. The results for ridges and rima were not so well defined. It was difficult to measure the width of the linear feature on the face of the monitor precisely. The width at the limits of recognition was on the order of 1/16 inch to 1/32 inch. As a result it was felt that at least 5 to 8 resolvable elements

across the ridge or rima were necessary to determine the bright side and the dark side of the feature.

2.3 Photographic to Digital Conversion

The procedure for converting the photographic pattern sets to pattern files stored on digital magnetic tapes in a form suitable for use in the simulation experiments included the following steps:

Developing a video image of the lunar photograph with a slow scan television system.

Recording the video output of the TV system on analog magnetic tape.

Replaying the recorded video signal through analog to digital conversion equipment creating a set of digitized video tapes.

Processing the digitized tapes to obtain correct resolution, enhance contrast, and eliminate "between picture" noise.

Editing the processed pictures to separate training and independent (generalization) samples and to provide translated versions of the recorded pictures.

The video images of the pattern photographs were obtained with a slow scan television system, Model 6030B, manufactured by General Electrodynamics Corporation of Garland, Texas. The specifications for this equipment are provided as Appendix A to this report. The slow scan camera was mounted vertically on a stable structure with the lens pointing down toward the center of a 4 foot by 4 foot table. The table was easily moved up and down providing a practical range from 3 inches to 50 inches between lens and photograph. A suitable set of close up lens attachments provided an image of satisfactory clarity over this entire range to enable increasing or reducing its size with respect to the photographic pattern. A photo flood light was mounted about 2 inches from the camera on each side of the camera to eliminate shadows cast by the camera and supporting structure. The intensity of the two lights was smoothly controlled by a Variac. The vidicon tube used in the slow scan camera integrated the incident light to produce an image requiring a semidarkened room for the recording process.

A certain amount of operating experience was necessary to determine the best combinations of light intensity and exposure time for producing

sharp images without saturating the vidicon. These combinations would change with distance between the lens and picture and in some cases would change with the gray scale level and texture of the photograph. A Simmons Omega repeating timer was used to control the exposure interval.

The results of the resolution study indicated that 25 lines of resolution should be adequate to allow classification of the craters. It was decided to make the largest crater no larger than 80% of the width of the aperture to allow room for various translations. Since the smallest craters were to be no smaller than two thirds the size of the largest craters, a 50 x 50 point picture gave at least 25 lines of resolution across the smallest crater. Squares delimiting the smallest and largest crater sizes were drawn with grease pencil on the face of the TV monitor. The size of the chosen aperture was 1.5 inches on the face of the monitor.

The pictures were recorded with four times the final 50 x 50 resolution giving an initial aperture of 200 x 200 points. In the processing program described later in this section the resolution was reduced by a factor of four by summing the gray scale values of four adjacent points on a line for four lines and dividing by 16 to obtain a numerical average of the 16 points. This process yielded the desired 50 x 50 resolution across the aperture and alleviated the effects of noise added to the video signal in the recording process. It also tended to compensate for slight misalignments of individual scan lines.

The height of the picture on the monitor scope was 2.47 inches. In order to obtain 200 lines across the 1.5 inch aperture, 330 lines were required for the entire picture. The aspect ratio or ratio of width to height was 1.10, indicating that 363 sample points should be taken on each line.

The Precision Instruments' 2100 recorder was used to record the video output of the TV camera. The 30-inch/sec speed with its frequency response from dc to 10 KC yielded the best compromise between video scan time (the longer the scan, the worse the picture) and the number of pictures stored on each reel of analog magnetic tape (100). With careful calibration of the recorder electronics it was possible to achieve the specified signal to noise ratio of 46 db. However, it is evident that the RMS power of the gray scale information carried on the video pulse was much smaller than the dc

power of the video pulse itself; therefore, the amplitude of the noise generated in the recording process was in the same range as the gray scale signal. Averaging 16 sample points to obtain one point at reduced resolution virtually eliminated the effects of this noise.

The recording process was straightforward. The photographic pattern was placed on the table under the camera at a distance which would place the feature within the limits of the aperture marked on the monitor. Craters were centered between the minimum size and maximum size limits. The photographs were oriented so the sun was from a specific direction. The lights were turned on for the proper exposure time, the recorder was turned on and allowed to stabilize at 30 ips, and one TV scan was initiated. When the scan was complete the recorder was turned off. A polaroid photograph was made of the TV monitor to retain a record of the actual area of the photograph scanned and the probable quality of the picture. Each photographic pattern was recorded in three rotations differing by 15 degrees.

When approximately three analog reels were filled in pictures they were played back and processed with an analog to digital converter at a Douglas data processing facility. The following calculations provided an estimate of the sampling frequency required when digitizing the patterns.

$$\text{Desired visible horizontal trace time} = \frac{\text{Number of horizontal samples}}{2 \times (\text{Video bandwidth})}$$

$$= \frac{363 \text{ Samples}}{(2)\ (10^4 \text{ cps})} = 18.1 \text{ ms}$$

The actual video bandwidth was 1 MC but the 10 KC filter on the input of the FM recorder gave an effective video bandwidth of 10 KC. The flyback time was measured at 1.1 MS so the total horizontal sweep time was 19.2 MS.

$$\text{Total frame time} = (\text{Number of lines})\ (\text{Total horizontal sweep time})$$

$$= (330 \text{ lines})\ (19.2 \text{ MS}) = 6.33 \text{ sec}$$

Allowance for 7% blanking and 75% registration probability normally considered in this calculation was not included. Tests at the beginning of the study indicated blanking was negligible. The registration problem was virtually eliminated by the average of four points in both directions since the minimum

spatial frequencies appearing in the final picture would cover the width of eight lines per cycle instead of two.

The sampling frequency is then:

$$\text{Sampling frequency} = \frac{\text{Number of samples per line}}{\text{Visible horizontal trace time}}$$

$$= \frac{362 \text{ Samples}}{18.1 \text{ MS}} = 20 \text{ KC}$$

The tape speed was 15 ips and the sampling frequency was 10,000 samples/sec. Since the tapes were being played at half speed this gave an effective sampling rate of the required 20,000 samples/sec. The recorded video signal was quantized to about 900 levels from the lowest to the highest signal level, obviously more than adequate to represent the eight gray levels. The A/D converter was turned on at the beginning of a set of pictures and turned off after 18 pictures had been processed and the output had been recorded on a digital magnetic tape. No attempt was made to eliminate the "between pictures" interval during the digitizing process.

The resulting magnetic tapes with the digitized video data were processed on Astropower's SDS 930 computer. The processing program read the tapes and separated the pictures from the "between pictures" interval. At the same time, it performed resolution reduction by obtaining the numerical average of blocks of 16 points as described above.

The program also performed a gray scale expansion about the features of interest. The gray scale range was determined by considering the highest video level in the pictures as white and the lowest level as black. The intervening range was divided equally between the six remaining gray levels. The floating gray scale was needed to compensate for the constant adjustments of the video gain required to maintain satisfactory performance of the TV system during the recording phase. It had the additional effect of enhancing the picture contrast in the regions of interest, serving as a mild form of signal conditioning.

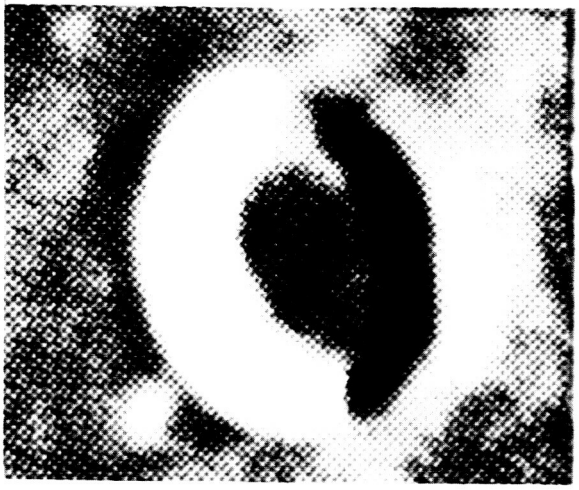
The result was an 85 x 72 matrix of numbers from 0 to 7 representing the brightness level at each point on the photograph. In order to interpret the processed pictures more easily a character printout was used

in addition to printing the numerical gray scale values. This consisted of assigning dense characters such as B's to represent dark points and letting blanks or periods represent light points. Intermediate points are represented by appropriate combinations of characters. The character printout was used for the reduced resolution examples shown in Figures 3 and 7.

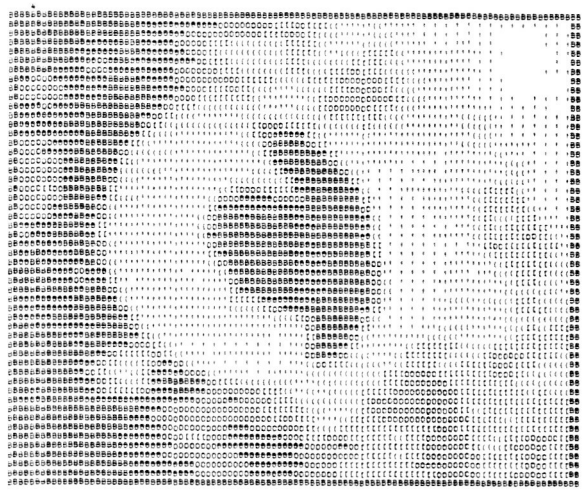
The final step in preparing the patterns was the editing process. Each of the 50 or so basic patterns was initially recorded in three rotations differing by about 15 degrees, giving us from 120 to 150 digital pictures for each set. The printouts were studied and the quality of each pattern was rated. The editing process consisted of choosing a 50 x 50 set of points from the larger 85 x 72 picture so that each feature was fairly well centered within its 50 x 50 frame. In addition, several translations were chosen which effectively shifted the feature within its frame up to 17 per cent of the width of field in four directions. In the case of the craters, nine translations were taken from each digitized picture. The linear features proved to have a wider variation in pattern quality, so up to 16 translations were taken from the better patterns. The editing resulted in a set of 1000 training and 200 independent test patterns for each of the four basic pattern sets.

Figures 3 through 6 show examples of each of the four basic patterns. The top two illustrations show the original photograph of a good example of the feature and its 50 x 50 character printout. The lower illustrations show other typical examples of the same feature. The illustrations underscore the fact that the distinguishing feature between the two types of crater patterns (the conspicuous central elevation) occupies only about 1 percent of the area of the picture. The distinguishing features between rima and ridges (the light and dark sides of the pattern) occupied from 5 percent to 10 percent of the picture area.

Figure 7 illustrates that the 50 x 50 point field preserves a surprising amount of detail found in the original picture. However, it is still difficult to determine from inspection of the reduced resolution picture that the crater does have a central elevation.

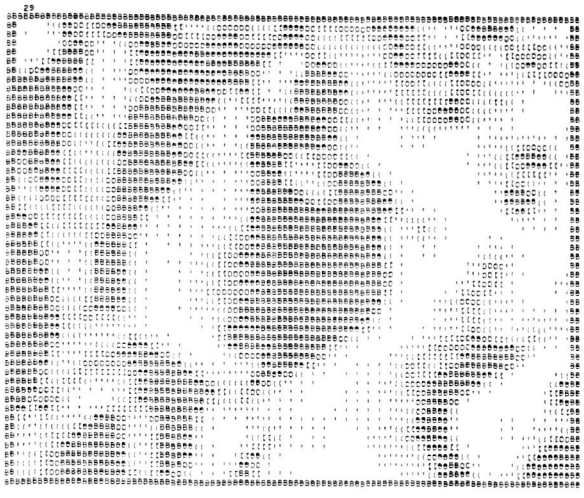


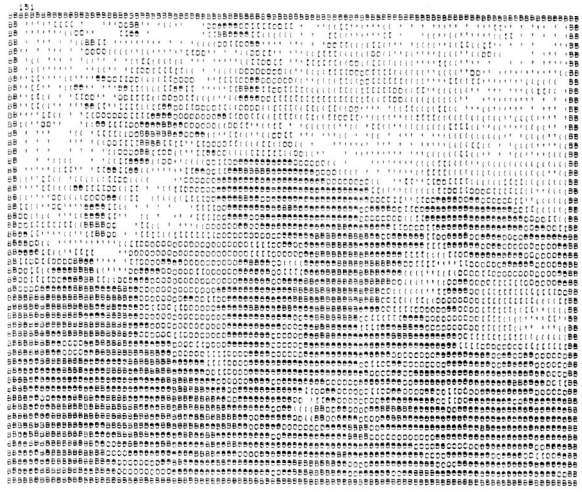
Original Photograph



Final Resolution

Good Crater Prototypes

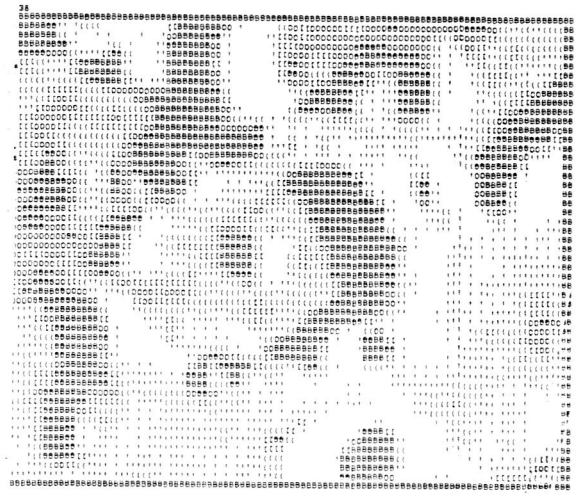




Typical Crater Patterns

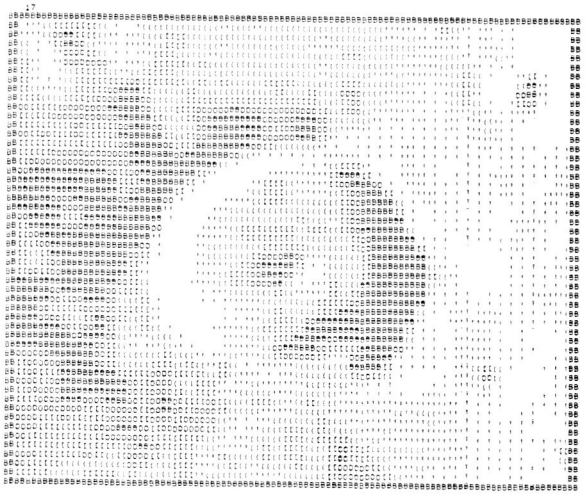
Figure 3. Lunar Craters With No Conspicuous Central Elevation

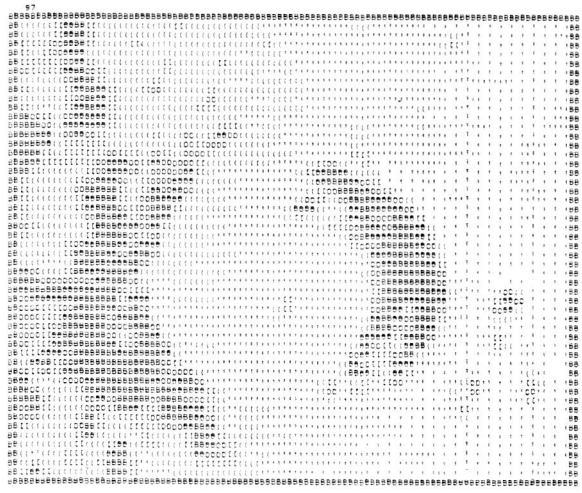
Original Photograph



Final Resolution

Good Prototype

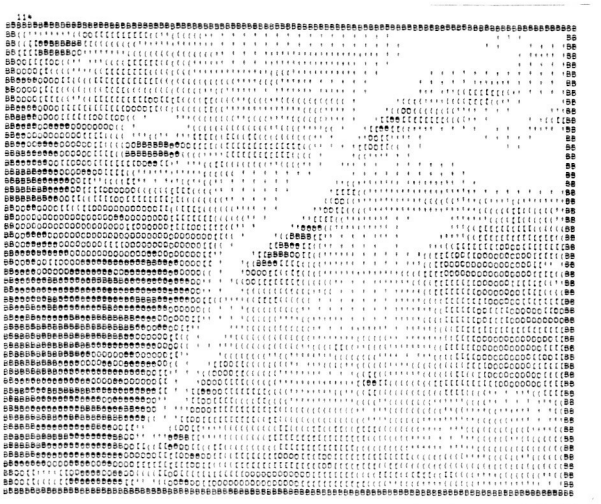




Typical Patterns

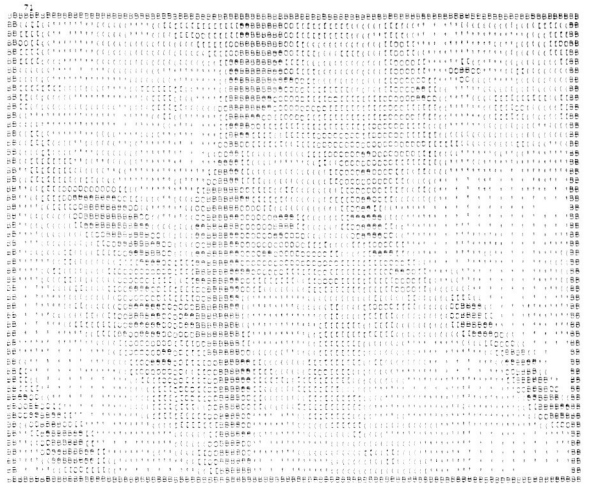
Figure 4. Lunar Craters With One or More Conspicuous Central Elevations

Original Photograph



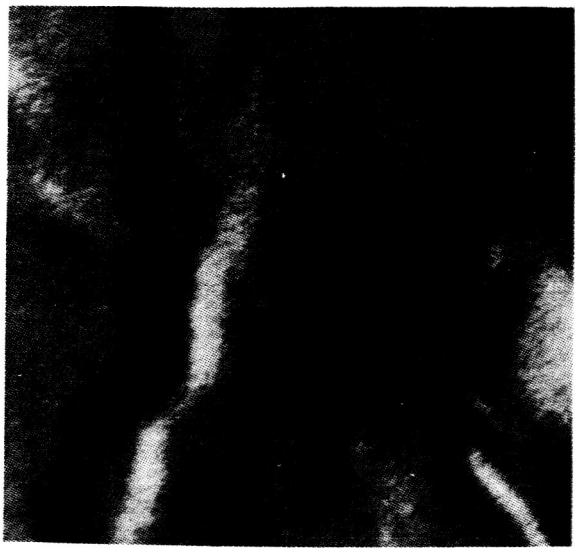
Final Resolution

Good Rima Prototype

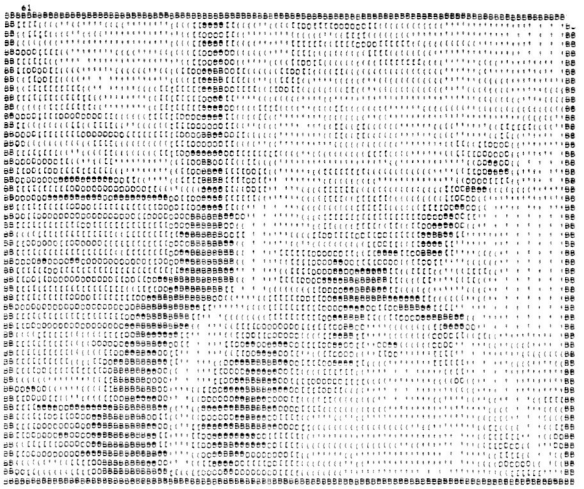


Typical Rima Patterns

Figure 5. Lunar Rima

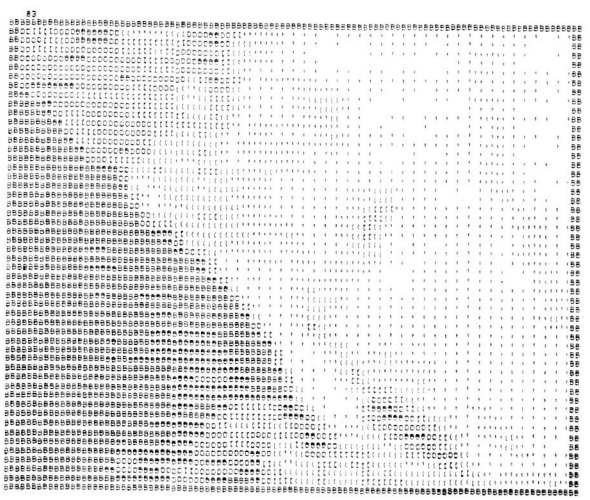


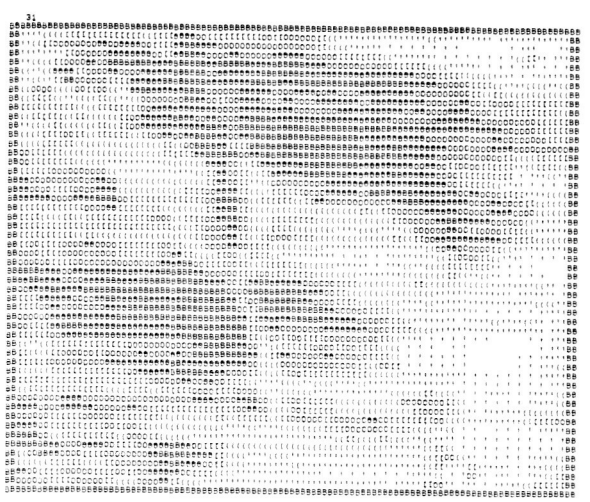
Original Photograph



Final Resolution

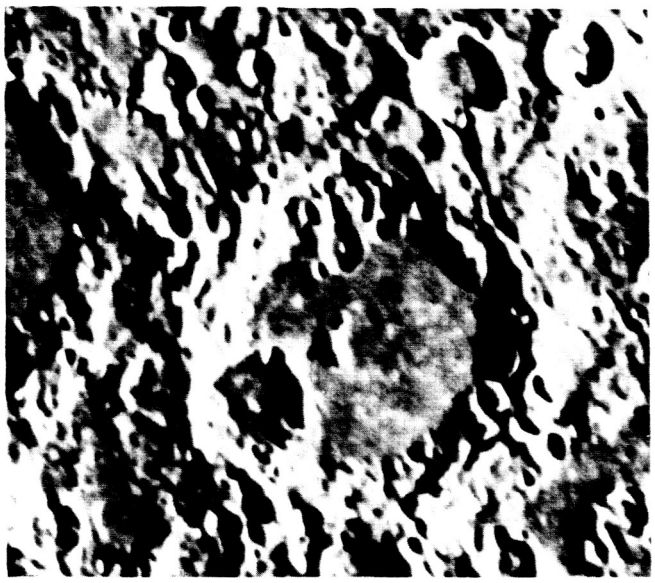
Good Ridge Prototype



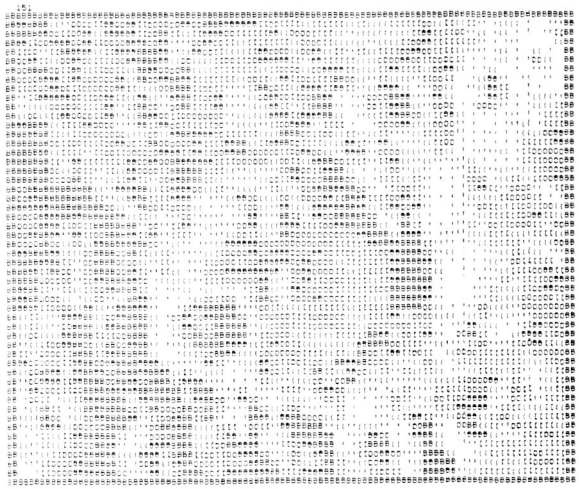


Typical Ridge Patterns

Figure 6. Lunar Ridges



Original Photograph



Final Resolution

Figure 7. Crater with Central Elevation (Albategnius)

3.0 DESIGN TECHNIQUES

3.1 General Discussion

The pattern recognition techniques under consideration share a common structure. As shown in Figure 8, properties are extracted from the (conditioned) pattern, and the decision function is computed from the property profile of the pattern.

In this study, two techniques were used to generate the property filters (frequently referred to as property detectors or logic units in the literature). Each property filter has a limited number of input connections and hence examines only a portion of the input data. With both techniques, the output of a property filter is binary. With one technique, the switching surfaces of the property filters are linear, and are based on the average brightness of the pattern classes at the appropriate points in the input field. With the other technique, the covariances of the brightnesses are considered, and the switching surfaces which result are quadratic. Detailed descriptions of the techniques are given in Section 3.3.

Six techniques for specifying the decision function were applied to the property profiles determined by the two methods mentioned above. Five of these result in linear decision functions (of the property profiles). In the sixth technique, a MADALINE system, the decision function is piecewise linear. Detailed descriptions of the six techniques are given in Section 3.2.

The two techniques for designing property filters were applied to three recognition tasks on the lunar data. For each of the six sets of property profiles generated a decision function was designed using each of the six techniques under consideration. The primary comparisons of Section 4 are based on the resulting 36 systems.

3.2 Decision Functions

The six techniques for designing decision functions are called, in this report:

1. Forced Learning
2. Bayes Weights
3. Error Correction

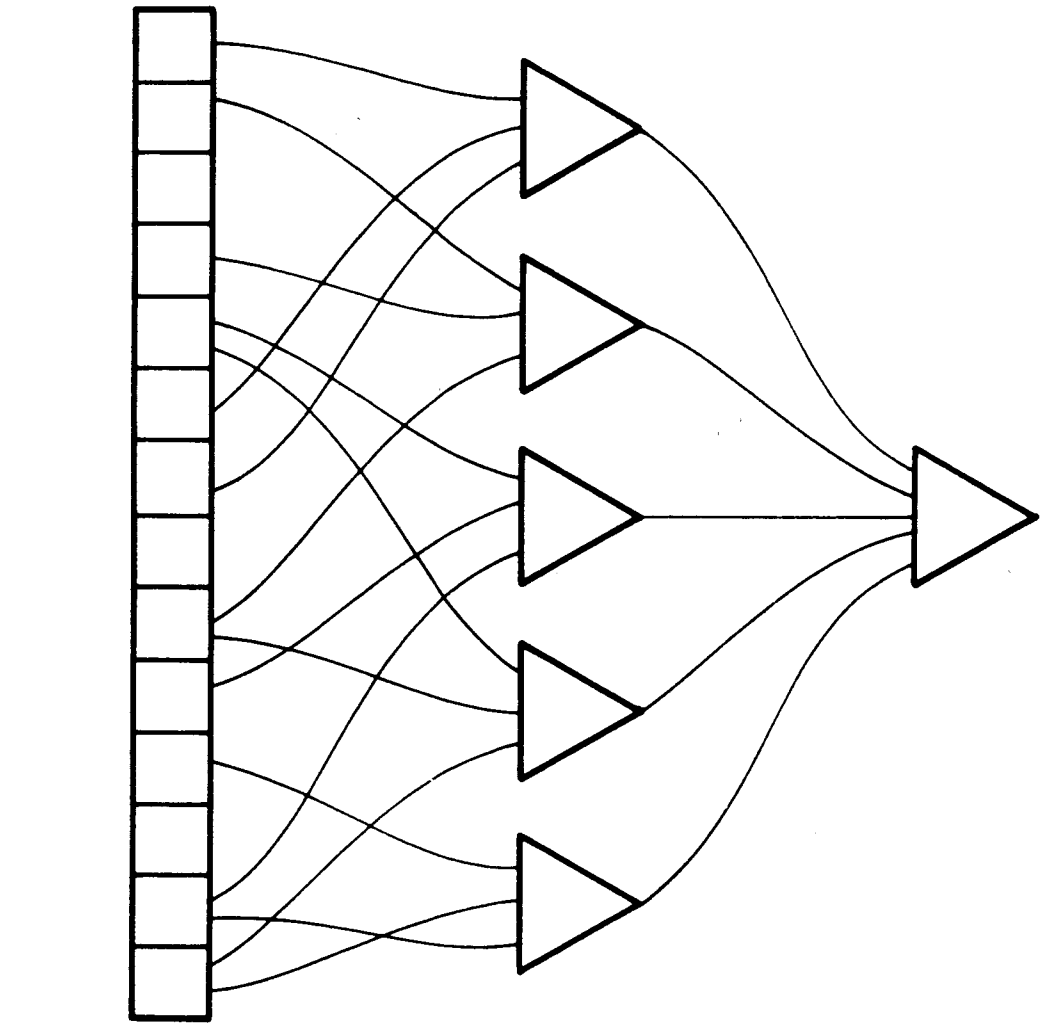


Figure 8. Decision Network Structure

4. Iterative Design
5. Mean Square Error
6. MADALINE

The first three of these are generally associated with the Perceptron program of F. Rosenblatt.[4] The last two are associated with the MADALINE program of B. Widrow.[5] Iterative design is a minimum loss approach, a technique first applied by W. Highleyman.[6] The particular algorithm used with its exponential form of the loss function, was developed at the Astropower Laboratories.

The first five techniques derive linear functions of the binary property profile vectors for the computation of the decision. Let $b_i(j)$ be the output of the i-th property filter for the j-th pattern. Let

$$D(j) = \Sigma w_i b_i(j) - \theta$$

where θ is a threshold for the decision element. Then a binary decision is achieved by assigning the j-th pattern to the "positive" class if $D(j) > 0$ and to the "negative" class if $D(j) < 0$. If $D(j) = 0$, it is assumed, in this study, that the decision is in error regardless of the actual classification of the j-th pattern. For computational convenience, the two values of $b_i(j)$ are taken to be "1" and "0" in the first four techniques, and "1" and "-1" in the last two techniques. This assumption does not affect the capabilities of the techniques. The first five techniques differ in the method for assigning values to the parameters w_i and θ, based on the property profiles of the sample patterns.

3.2.1 Forced Learning

In the forced learning technique, the oldest of the six methods, the weights w_i are defined to be:

$$w_i = p_i - q_i$$

where p_i is the fraction of patterns in the "positive" class possessing the i-th property (i.e., the fraction for which $b_i(j) = 1$) and q_i is the fraction of patterns in the "negative" class possessing the property (again the fraction for which $b_i(j) = 1$). If the number of samples for each pattern class is equal, as in this study, the weights can be derived adaptively. Let the number of samples per class be "m". Set each weight initially to zero. Then the patterns are

examined one by one. The weights are changed according to the following rule.

$$\Delta w_i(j) = \frac{1}{m} \text{sgn}(j)\, b_i(j)$$

where sgn(j) is "1" or "-1" according to whether the j-th pattern is "positive" or "negative" and $\Delta w_i(j)$ is the increment to the w_i due to the j-th pattern. Summary Figure 9 includes the forced learning technique.

The forced learning technique generally assumes the value of the threshold, θ, to be zero. In this study, it was taken to be the average value of

$$\Sigma w_i b_i(j)$$

for all sample patterns. Using this average value compensates for property filters which are designed to be "on" primarily for patterns of one class and "off" primarily for patterns of the other class. The original forced learning technique was applied to randomly generated property filters which did not have this consistent bias.

3.2.2 Bayes Weights

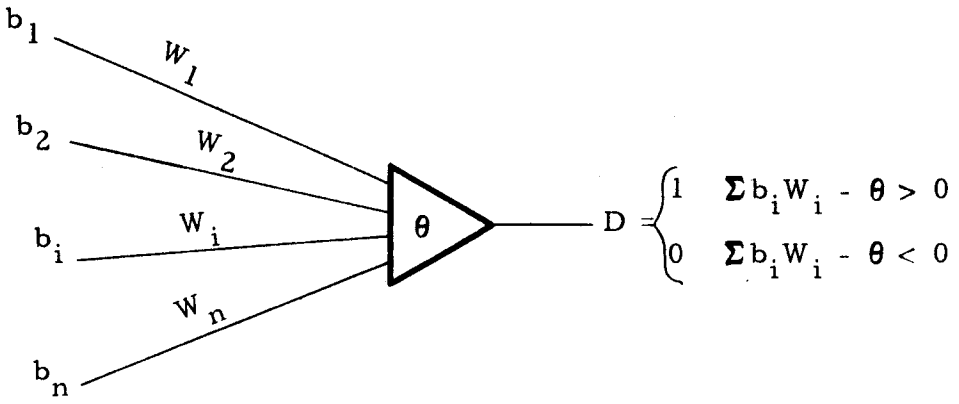
The Bayes weights technique was originally derived as an "optimum" solution to the assignment of a decision function under the assumption that the property filters were statistically independent. This assumption is not, in general, valid. With this technique, the weights are assigned by the following rule:

$$w_i = \ln \frac{p_i}{1 - p_i} - \ln \frac{q_i}{1 - q_i}$$

where the symbols are as described above.* These weights are not derived adaptively, but may be computed from the forced learning data if the accumulations of $\Delta w_i(j)$ are kept separately for "positive" and "negative" patterns. As with the forced learning, the threshold θ is taken as the average value of $\Sigma w_i b_i(j)$. The Bayes weights technique is summarized in Figure 9.

---

*To prevent the occurrence of undesirable infinite weights, a small constant is added to each of the four probability estimates p_i, $1 - p_i$, and $1 - q_i$. This may be justified by a Bayes argument in which an a priori Beta distribution is assumed (Reference 8).



Forced Learning

$$W_i = p_i - q_i$$

Adaptively $\Delta W_i(j) = \frac{1}{m} (sgn) (b_i(j))$

Bayes Weights

$$W_i = \ln \frac{p_i}{1 - p_i} - \ln \frac{q_i}{1 - q_i}$$

Error Correction

Adaptively $\Delta W_i(j) = (err) (sgn) (b_i(j))$

Mean Square Error

Adaptively $\Delta W_i(j) = \frac{1}{n+1} (1 - (sgn) (\Sigma_i b_i(j) W_i - \theta)) (b_i(j))$

C/829

Figure 9. Decision Function Generation

3.2.3 Error Correction

The error correction procedure is largely responsible for the current popularity of adaptive pattern recognition. The error correction theorem guaranteed perfect performance on the sample patterns, provided that a set of weights permitting this existed.

The weights are derived adaptively, in a manner similar to the adaptive derivation of the forced learning weights:

$$\Delta w_i(j) = \text{sgn}(j)\, b_i(j)$$

The difference is that this weight change is applied only when the j-th sample pattern is misclassified. If the sample pattern is correctly classified, the weights are not changed. This modification requires that each sample pattern be processed many times. The error correction theorem specifies that only a finite number of corrections are needed to achieve a solution, but does not specify the number. The patterns were processed cyclically in this study.

The threshold θ is also determined adaptively.

$$\Delta\theta(j) = -\text{sgn}(j)$$

Again, this change is made only when an incorrect decision is observed.

A modification of the original algorithm is incorporated to speed convergence. Whenever the length of the coefficient vector (the weights and the threshold) exceeds its previous maximum, all of the coefficients are doubled. This change was verified in a number of experimental runs, and is based on previous experiments and analytic considerations.

The error correction technique is summarized in Figure 9.

3.2.4 Mean Square Error

The mean square error technique differs from the error correction method in that, for design purposes, it is assumed that the desired value of

$$\Sigma w_i b_i(j) - \theta$$

is "1" for "positive" patterns and "-1" for "negative" patterns. Each time a

pattern is processed, the weights and threshold are changed to give this desired state. This is facilitated by using the values of "1" and "-1" for $b_i(j)$.

$$\Delta w_i(j) = \frac{b_i(j)}{n+1} \left(\text{sgn}(j) - (\Sigma w_i b_i(j) - \theta) \right)$$

where n is the number of property filters. The threshold is derived similarly,

$$\Delta\theta(j) = -\frac{1}{n+1} \left(\text{sgn}(j) - (\Sigma w_i b_i(j) - \theta) \right)$$

Unlike error correction $\Delta w_i(j)$ can be positive or negative for a pattern of either class, and may change sign for one pattern processed at different times.

The patterns are processed cyclically. The technique is summarized in Figure 9.

3.2.5 Iterative Design

The iterative design technique is based on a minimum loss approach. Each sample pattern is assigned a loss number which depends on the value of the decision function.

$$L_j = \exp \left\{ -\text{sgn}(j) \left(\Sigma w_i b_i(j) - \theta \right) \right\}$$

The loss for the j-th pattern, L_j, will be less than "1" if the pattern is correctly classified and greater than "1" if it is incorrectly classified. The magnitude of the loss number reflects how definite is the decision. A system loss is defined as the sum of the individual pattern losses.

$$SL = \Sigma L_j$$

The weights and threshold are adjusted to minimize the system loss. Since this cannot be done directly, a relaxation process is used. The weights are adjusted cyclically. Each time a weight is changed, the threshold is also adjusted. These changes are defined by

$$\Delta\theta = \frac{1}{2} \ln \frac{\sum_j L_j (1 - b_i(j)) (1 - \text{sgn}(j))}{\sum_j L_j (1 - b_i(j)) (1 + \text{sgn}(j))}$$

and

$$\Delta w_i = \Delta\theta - \frac{1}{2} \ln \frac{\sum_j L_j b_i(j) (1 - \text{sgn}(j))}{\sum_j L_j b_i(j) (1 + \text{sgn}(j))}$$

It has been shown that this process will give system losses converging to a minimum value for the system loss. In common with the error correction technique, perfect performance on the sample patterns will be achieved whenever this is possible.

To prevent the possibility of the finite samples yielding infinite weights (and hence absolute indicators) when not warranted, a small constant is added to each of the summations shown above.

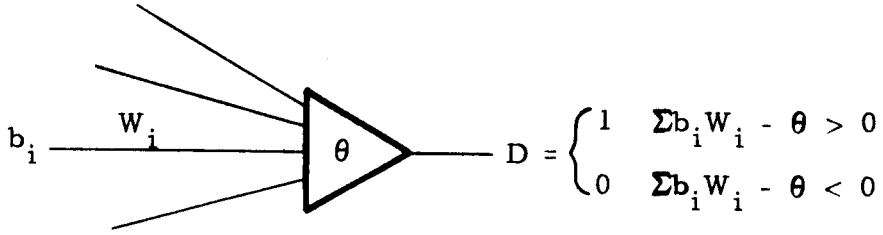
The iterative design technique processes one property filter at a time, but needs data from all of the sample patterns. The computer requires "unit profiles" for the sample patterns, rather than property profiles. The technique is summarized in Figure 10.

3.2.6 MADALINE

The MADALINE technique is the only one considered capable of producing a nonlinear decision function. It does so at the expense of system complexity (see Figure 11).

The design technique has the added job of allocating the task among the subdecision elements, or ADALINES. The algorithm selected is an extension of the error correction technique.

The basic principle behind the algorithm is that of minimum change. As a sample pattern is processed, the first step is to determine if the decision is correct. If it is, no changes are made. If it is not, the number of ADALINES whose decisions must be changed to give a correct majority vote is noted. This number of ADALINES is selected from the (incorrect) majority on the basis of minimum magnitude of the linear subdecision functions. Each of the selected ADALINES is given a sufficient number of corrections, according to the error correction technique, to reverse its decision. It is to be noted that the minimum change principle is simplified by using "1" and "-1" values for $b_i(j)$.



Iterative Design

$$L_j = \exp\left\{(sgn)\ (\theta - \sum_i b_i(j)\ W_i\right\}$$

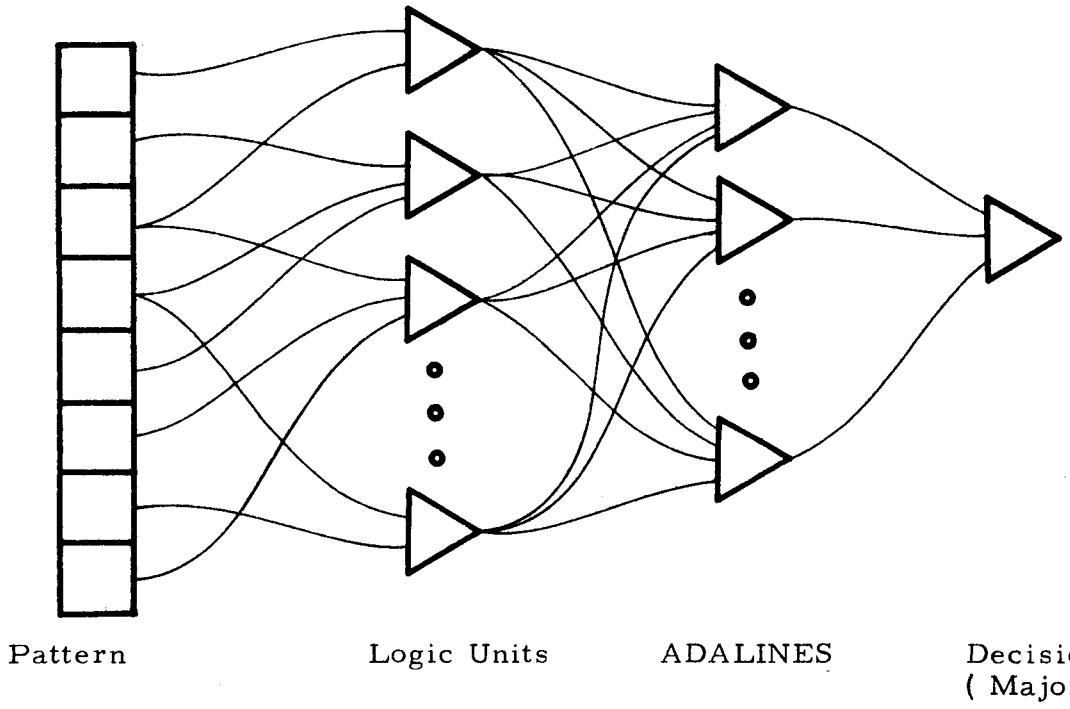
$$SL = \sum_i L_j$$

Adaptively $\Delta\theta = \frac{1}{2} \ln \frac{\sum_{neg} L_j}{\sum_{pos} L_j}$ Sums for $b_i = 0$

$$\Delta W_i = \Delta\theta - \frac{1}{2} \ln \frac{\sum_{neg} L_i}{\sum_{pos} L_i}$$ Sums for $b_i = 1$

c/828

Figure 10. Decision Function Generation



C/827

Figure 11. Madaline Structure

The patterns are processed cyclically. Convergence to a solution, when it exists, is not guaranteed, and indeed, counterexamples to convergence have been found. Figure 12 summarizes the technique.

3.3 Property Filters

Two techniques for generating binary property filters, or logic units, are used in this study, one yielding units with linear switching surfaces and one yielding quadratic input units. The units have a limited set of randomly selected input connections.

Each logic unit thus views the patterns as projected on a subspace of the signal space. The switching surface for the unit is derived from the analysis of the moments of the sample patterns as projected into the subspace. The analyses are optimum if the underlying pattern distributions are multivariate Gaussian. The linear surface arises as a special case when the covariance matrices for the pattern classes are assumed equal, and a multiple of the identity matrix.

3.3.1 Linear Units

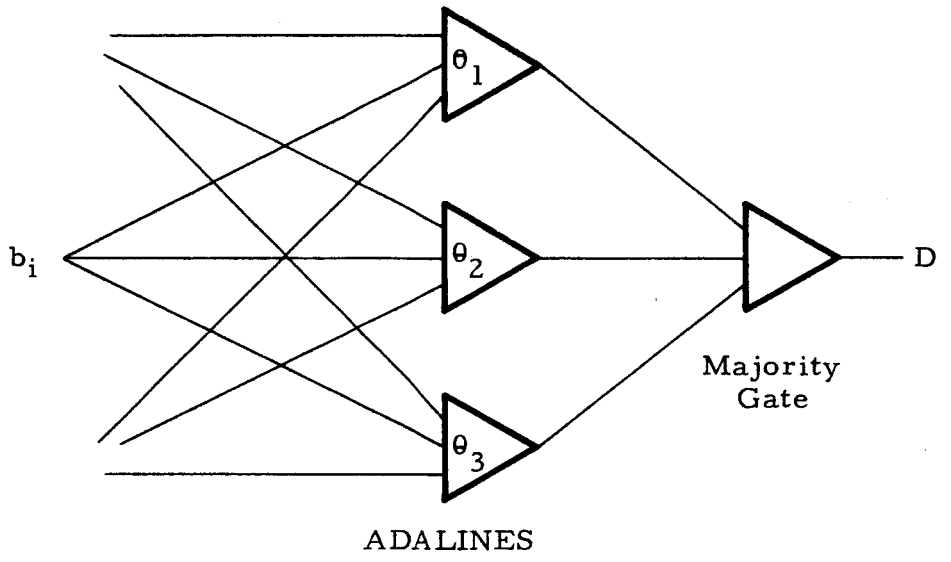
For each of the recognition tasks, a set of 1000 property filters was derived. Each unit has 10 randomly selected input connections. Within the 10-dimensional subspace, the logic unit is defined by the vector equation:

$$X^T(M_1 - M_2) - \frac{1}{2}(M_1^T M_1 - M_2^T M_2) = 0$$

where M_1 and M_2 are the sample mean vectors of the "positive" and "negative" classes within the subspace. The switching surface consists of vectors X which satisfy this equation.

As shown in Figure 13, the switching surface consists of a hyperplane which is the perpendicular bisector of the line segment joining the means M_1 and M_2. The equation of the switching surface might also have been written:

$$\Sigma a_i x_i - \theta = 0$$



C1826

1. Test Decision
2. Determine Number of ADALINES to be Changed if Decision Incorrect
3. Choose those ADALINES Requiring Least Change

Figure 12. Madaline Technique

The coefficients a_i are given by

$$a_i = (m_1)_i - (m_2)_i$$

that is, the difference in the mean brightness for the two pattern classes at the appropriate sensory field point. It is to be noted that if the patterns were binary valued these coefficients would be identical to those derived from the property profiles by the forced learning technique described earlier.

The advantage and the weakness of the technique are apparent. The coefficient for a connection does not depend on which other connections are selected for the unit, thus the units are very easily generated. However, the units are not designed based on contrasts and gradients in the patterns, and require a high degree of pattern centering, size normalization, and brightness and contrast control if the units are to be effective.

If the discussions of Section 4, the linear units are referred to by the code "SDA" for simple discriminant analysis.

3.3.2 Quadratic Units

For each recognition task, a set of 310 quadratic logic units, each with seven input connections, was designed. Within a randomly selected subspace, the switching surface of a unit is defined by

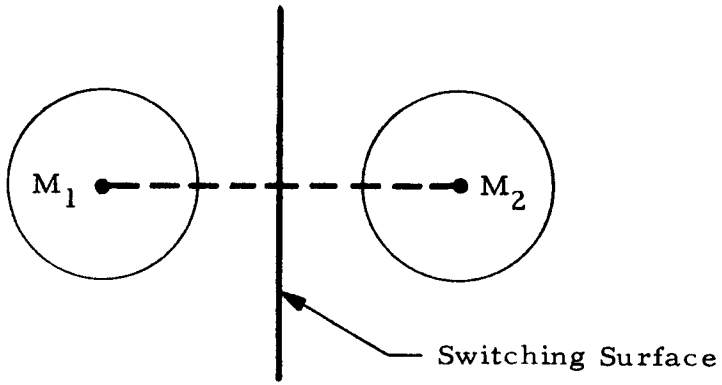
$$X^T(\Sigma_1^{-1} - \Sigma_2^{-1})\,X - 2X^T(\Sigma_1^{-1}M_1 - \Sigma_2^{-1}M_2) + M_1^T\Sigma_1^{-1}M_1 - M_2^T\Sigma^{-1}M_2 + \ln\frac{|\Sigma_1|}{|\Sigma_2|} = 0$$

The vectors M_1 and M_2 are the sample mean vectors, as above, and the matrices Σ_1 and Σ_2 are the sample covariance matrices of the "positive" and "negative" classes within the subspace. The switching surface is illustrated in Figure 13.

Since the equation of the switching surface does incorporate the covariance matrices, the resulting logic units do reflect average contrasts and gradients for pairs of sensory field points. The price for this is the necessity for matrix inversion, and the complexity of the quadratic input property filters.

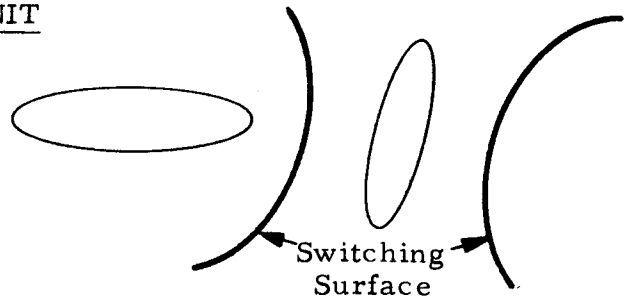
The quadratic units were designed by a simulation program in which a decision function is also defined (using the iterative design

SIMPLE DISCRIMINANT ANALYSIS



$$X^T (M_1 - M_2) - \frac{1}{2} (M_1^T M_1 - M_2^T M_2) = 0$$

QUADRATIC UNIT



$$X^T(\Sigma_1^{-1} - \Sigma_2^{-1}) X - 2X^T(\Sigma_1^{-1}M_1 - \Sigma_2^{-1}M_2) + M_1^T\Sigma_1^{-1}M_1 - M_2^T\Sigma_2^{-1}M_2 + \ln\frac{|\Sigma_1|}{|\Sigma_2|} = 0$$

Figure 13. Property Filter Generation

technique). The process is as follows. 100 subspaces are selected and quadratic units defined. The units are then evaluated on the basis which unit would result in the lowest system loss (SL) if considered in a one property filter network. The 10 best units are retained, and the remaining 90 are discarded. The output weights are assigned as is the decision threshold in a single pass through the weights. 100 new candidate units are then generated. Holding the existing weights fixed, the candidates are evaluated as individual additions to the existing network. The 10 units resulting in the greatest reduction in the SL are retained and the remaining 90 discarded. A single pass through the weights with the relaxation process is used to update the linear decision function. The process is continued in this fashion, each candidate unit being evaluated as an addition to the existing network. Due to computer memory restrictions, each logic unit output weight is changed only twice after the initial assignment before it is frozen.

Once the network is designed, the decision function is discarded, and the property filter set is retained for use in the study. These units are designated in Section 4 by the code "DAID" for discriminant analysis-iterative design. In one case, the set of property filters used is the combined initial segments of two network designs. Performance levels achieved by subsequently assigned decision functions exceeded those of the original network, due to the limitation placed on the number of weight adjustments in the DAID process.

4.0 EXPERIMENTAL PROGRAM

4.1 Introduction

Twelve combinations of design techniques were tested on three recognition tasks dealing with lunar topography. The experiments are summarized in Figure 14.

Of the three tasks, separating the craters with central elevations from those without is the most difficult. Although the craters themselves are usually well defined, the central elevations are not. The elevations, which are the significant portion of the image in this task, cover less than one percent of the sensory field. Separating the ridges from the rima represents a task of intermediate difficulty. The patterns in general are not well defined. Inspection of the computer printed patterns reveals a number of samples which are not readily classified by the observer. The significant portions of the patterns are larger than in the crater problem, covering perhaps five to fifteen percent of the sensory field. The task of separating the craters from the linear features is the easiest of the three, as the significant feature, the crater, is usually well defined and may constitute more than thirty percent of the image.

For each of the three tasks, two sets of property filters were designed as discussed in Section 3.3. One set consisted of 1000 linear units, each having 10 input connections. This set is designated by the code "SDA." Each set of linear units required three hours of time on an SDS 930 computer to generate the units and to determine the property profiles of the sample patterns. The other set consisted of 310 quadratic units, each having 7 input connections. The quadratic units are designated by the code "DAID." Generation of the units and the corresponding property profiles required seven hours of computer time on an IBM 7094. The design was accomplished using 1000 sample patterns for each class. A completely independent sample of 200 patterns for each class was used to test the generality of the network designs.

For each of the three tasks, and for each of the two sets of property filters, six techniques were used to design decision functions. The six techniques are also discussed in Section 3.0. All of these functions were determined on the SDS 930. The first two methods, forced learning and Bayes weights, were combined in a single program which had a running time of 20

TASKS

1. Craters with Central Elevations vs. Craters Without
2. Rima vs. Wrinkle Ridges
3. Craters vs. Linear Features

PROPERTY FILTER GENERATION

1. Simple Discriminant Analysis (SDA)
2. Discriminant Analysis with Iterative Design (DAID)

DECISION FUNCTION GENERATION

1. Forced Learning
2. Bayes Weights
3. Error Correction
4. Iterative Design
5. Mean Square Error
6. MADALINE

c/824

Figure 14. Experimental Program (Lunar Features)

minutes. The remaining four methods are recursive. Each of these methods had its own computer program. Each program was allowed to run a minimum of eight hours (maximum 12 hours) for each decision function, the exact time depending on computer availability. Three exceptions occurred when the programs terminated upon reaching perfect performance. More cycles could be accomplished on the DAID units than on the SDA units, since there were fewer adjustments to be made in each cycle.

4.2 Preliminary Experimentation

The computer programs were available before the digitized sample patterns were completed. To test the programs, sets of patterns with two bit gray scales were generated randomly. Based on experiments performed on these patterns, using SDA units, a number of parameters were adjusted to the design algorithms.

The crater patterns were obtained two months before the linear feature samples were completed, and the experimentation was continued on the crater separation problem. Earlier changes were confirmed, and additional adjustments were made, based on the results of these experiments.

Most of these adjustments appear to have improved the programs. In one case, a change proved to be detrimental to the error correction method used. This change will be removed, and the appropriate programs will be rerun during the coming reporting period. A parameter choice for MADALINE does not appear to be a good one and will be modified. Five of the major decisions will be mentioned here.

First, it was noted that on the random patterns, error correction produced unbalanced designs. Nearly all of the incorrect decisions would occur in one pattern class. To remedy this, at the end of the design program the decision threshold designed by the algorithm was discarded, and a new one determined as in the forced learning and Bayes weights techniques (the threshold is taken to be the average value of the decision function for all sample patterns). This not only eliminated the unbalance, but actually led to fewer classification errors on the random patterns. The results on the crater separation task were less definite – performance was unchanged by the substitution. Later results obtained on the remaining two tasks indicate that the substitution is undesirable. There was not sufficient time in the period to undo the change, however.

Based on the results obtained on the random data, the number of ADALINES used in the MADALINE system was set at seven (the maximum permitted by core limitations). In the experiments on the randomly generated data, the added structure permitted perfect classification of the sample patterns. When the results were disappointing on the crater vs. linear feature task, the program was rerun using three ADALINES. This permitted more cycles to be accomplished, and reduced the task of distributing the work among the ADALINES. Better performance was achieved. In the coming experiments on NIMBUS data, three ADALINES and longer design runs will be utilized.

A routine which doubles the weights and threshold whenever the length of the coefficient vector exceeds its previous maximum was added to the error correction program. With doubling, there is about an 80 percent reduction in the number of cycles needed to achieve a given performance level on the random patterns.

Experiments on the craters with the error correction technique indicated that the number of classification errors on the sample patterns oscillates violently from one cycle to the next. A ratio of 3 to 1 has been observed in consecutive cycles. Limited experiments indicated that the generalization performance followed the classification performance. For this reason, the error correction program was run twice on each problem – once to determine at which cycle the best classification performance was achieved, and the second time to terminate the design at that point. In the coming NIMBUS experiments, classification and generalization performance will be tested at the end of each cycle for all four recursive techniques.

Three of the techniques are sensitive to the sequence in which the patterns are processed. These are the error correction method, mean square error, and the MADALINE technique. The sequence is set so that it alternates between "positive" and "negative" patterns. Within the "positive" and "negative" subsequences, no special arrangements were made initially. Since a given crater occurs in three rotations and nine translations, the subsequences contained strings of 27 different views of the same crater. This proved to be detrimental, and the subsequences were rearranged so that no two consecutive patterns contained the same crater. The performance of the three systems averaged 11.4 percent higher classification rates and 4.8 percent higher generalization rates

when designed on the scrambled sequence. The sequence was scrambled for the remaining tasks.

4.3 Recognition Experiments

4.3.1 Craters vs. Craters

The task of separating the craters with central elevations from those without is the most difficult of the three tasks. The results obtained with the linear SDA units are given in Figure 15, and those with the quadratic DAID units in Figure 16. In these figures, the first column lists the technique and the second column the number of cycles for the recursive process. The third set of columns gives the percentages of correct decisions achieved on the sample patterns. A percentage is given for each pattern class as well as the overall figure. The next set of columns gives similar data for the generalization patterns. It is to be noted that there are two sets of classification figures listed for error correction. The bottom set represents performance with the threshold designed by the algorithm, the upper set with the substitute threshold discussed in Section 4.2. Generalization was tested against this latter threshold. The final set of data lists generalization performance when the decision threshold is adjusted to give maximum generalization performance. These figures are useful in assessing the effect of the threshold substitution.

Figure 17 summarizes Figures 15 and 16, presenting only the overall performance data. A distinct bias in favor of the DAID units may be observed in this figure. With these units, MADALINE is able to achieve perfect classification performance, error correction and iterative design achieve an intermediate level (92-93%), and the remaining three techniques give the poorest performance (82-84%). Bayes weights gives the best generalization performance, closely followed by forced learning. Mean square error is poorest in this department, while the other techniques are approximately equal. The average threshold used in the forced learning, Bayes weights, and error correction techniques achieves nearly maximal results with the DAID units, but falls down badly with the SDA units. The performance figures achieved on this difficult task are not satisfactory for a preliminary design, so that the comparisons drawn from them should not be given much emphasis.

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		With	W/O	Tot.	With	W/O	Tot.	With	W/O	Tot.
Forced L.		67.0	63.9	65.45	43.5	57.5	51.00	25.0	99.5	62.50
Bayes W.		67.0	64.0	65.50	43.5	57.5	51.00	25.0	99.5	62.50
Error Corr.	99	82.9	91.2	87.05	49.0	58.0	53.50	77.0	45.0	61.00
		94.2	80.0	87.10						
Iter. Des.	30	90.0	89.6	89.80	52.0	58.0	55.00	71.5	47.5	59.50
Mean Sq. Er.	24	55.7	59.0	57.35	53.0	51.0	52.00	85.5	29.0	57.25
Madaline	30	72.1	72.3	72.20	52.5	55.5	54.00			

Figure 15. Craters With Central Elevations vs. Craters Without 1000 SDA Logic Units

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		With	W/O	Tot.	With	W/O	Tot.	With	W/O	Tot.
Forced L.		84.9	80.3	82.60	64.5	61.0	62.75	90.0	37.5	63.75
Bayes W.		87.9	80.3	84.10	67.0	60.5	63.75	73.5	57.5	65.50
Error Corr.	63	94.0	91.6	92.80	71.5	46.0	58.75	61.5	61.5	61.50
		94.2	91.5	92.85						
Iter. Des.	60	92.0	92.2	92.10	59.5	58.5	59.00	80.0	43.0	61.50
Mean Sq. Er.	60	76.3	87.3	81.80	40.0	66.0	53.00	69.0	43.0	56.00
Madaline	32	100	100	100	59.5	56.5	58.00			

c1815

Figure 16. Craters With Central Elevations vs. Craters Without 310 DAID Logic Units

Technique	% Design Pats.		% Gen. Pats.		Best % Gen. Pats.	
	SDA	DAID	SDA	DAID	SDA	DAID
Forced L.	65.45	82.60	51.00	62.75	62.50	63.75
Bayes W.	65.60	84.10	51.00	63.75	62.50	65.50
Error Corr.	87.05	92.80	53.50	58.75	61.00	61.50
	87.10	92.85				
Iter. Des.	89.80	92.10	55.00	59.00	59.50	61.50
Mean Sq.Er.	57.35	81.80	52.00	53.00	57.25	56.00
Madaline	72.20	100.00	54.00	58.00		

Figure 17. Craters With Central Elevations vs. Craters Without

4.3.2 Ridges vs. Rima

The separation of the rima from the wrinkle ridges is the task of intermediate difficulty. Figures 18 and 19 present the results obtained with SDA and DAID units respectively, and Figure 20 summarizes the results.

These results again show a distinct advantage for the quadratic DAID units over the linear SDA units. With the DAID units, both error correction and iterative design are able to achieve complete separation of the sample patterns. When the threshold substitution was made in the error correction design, performance dropped markedly. In Figure 19, the two values in the cycles column for iterative design indicate that complete separation was obtained in 28 cycles, but the design process was allowed to continue for an additional 5 cycles. MADALINE, despite having seven times as much structure, did not achieve complete separation of the sample patterns. It is felt that this is attributable to the inefficient algorithm which could not arrive at a suitable state within the allotted time. A longer design run, and/or the use of fewer ADALINES might have yielded perfect separation.

The highest generalization performances are obtained by MADALINE and Bayes weights using the DAID units, but the levels achieved by all systems (except error correction) do not appear to be significantly different. Only the Bayes weights and forced learning techniques approach this level using the SDA units.

The exceptional case of error correction appears to be due to the threshold substitution when high classification performances are achieved (see also the crater vs. linear feature results, Section 4.3.3). The results for the best performance threshold (Best % Gen. Pats. column) seem to be comparable with the results with the other techniques. In the other five of six cases with forced learning, Bayes weights, and error correction, the average value threshold provides near maximal generalization performance.

An examination of the patterns shows a number of patterns for which the classification by an observer cannot be made accurately, and some patterns in which the ridge or rima is difficult to locate. In view of the

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		Rima	Ridges	Tot.	Rima	Ridges	Tot.	Rima	Ridges	Tot.
Forced L.		66.3	74.5	70.40	79.5	65.0	72.25	91.5	59.0	75.25
Bayes W.		66.3	74.5	70.40	80.0	65.0	72.50	91.5	59.0	75.25
Error Corr.	99	83.9	96.6	90.25	47.0	77.0	62.00	75.0	59.0	67.00
		93.4	90.6	92.00						
Iter. Des.	24	90.9	93.3	92.10	67.0	61.0	64.00	78.0	53.5	65.75
Mean Sq.Er.	30	77.8	79.3	78.55	53.5	59.0	56.25	43.5	71.5	57.50
Madaline	25	80.5	80.1	0.30	75.5	59.5	67.50			

c/817

Figure 18. Rima vs. Wrinkle Ridges (1000 SDA Logic Units)

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		Rima	Ridges	Tot.	Rima	Ridges	Tot.	Rima	Ridges	Tot.
Forced L.		90.3	72.1	81.2	73.5	70.5	71.0	59.5	89.0	74.25
Bayes W.		92.2	73.0	82.6	73.5	77.0	75.25	65.0	92.0	78.50
Error Corr.	190	100	86.9	93.45	100	2.0	51.0	63.5	85.0	74.25
		100	100	100						
Iter. Des.	33/28	100	100	100	78.5	69.0	73.75	67.0	85.0	76.00
Mean Sq. Er.	60	88.0	94.7	91.35	84.0	65.0	74.50	81.5	71.5	76.50
Madaline	50	96.1	95.5	95.80	80.0	71.5	75.75			

c/B/B

Figure 19. Rima vs. Wrinkle Ridges (310 DAID Logic Units)

Technique	% Design Pats.		% Gen. Pats.		Best % Gen. Pats.	
	SDA	DAID	SDA	DAID	SDA	DAID
Forced L.	70.40	81.20	72.25	71.0	75.25	74.25
Bayes W.	70.40	82.60	72.50	75.25	75.25	78.50
Error Corr.	90.25	83.45	62.00	51.0	67.00	74.25
	92.00	100.00				
Iter. Des.	92.10	100.00	64.00	73.75	65.75	76.00
Mean Sq.Er.	78.55	91.35	56.25	74.50	57.50	76.50
Madaline	80.30	95.80	67.50	75.75		

Figure 20. Rima vs. Wrinkle Ridges

quality of the patterns, generalization performance of 76 percent is encouraging, but it is not high enough for an operational system.

4.3.3 Craters vs. Linear Features

The separation of the craters from the linear features is the easiest of the three tasks. Despite this, it is the only task for which none of the systems achieved perfect separation of the sample patterns. The results on this task are given in Figures 21, 22, and 23.

In most cases the DAID units again proved to be more suitable than the SDA units. With the DAID units, iterative design and error correction give the best classification performance, yielding over 99 percent correct decisions. Even with the SDA units, these two techniques achieved higher classification performances than the other techniques do with either set of units. Since error correction classification performance is high on both sets of units, the effects of the threshold substitution are observable twice.

Generalization performance with the iterative design decision function equals the classification performance at 99.5 percent, considerably higher than the other techniques. It is likely that when recomputed without the threshold substitution, generalization performance with error correction will also be satisfactory.

With the SDA units, generalization performances are comparable to those achieved on the rima vs. ridges problem with the DAID units. As in that case, all techniques provide about the same level of performance.

Again, the MADALINE performances were disappointing. To test whether this was due to inefficiency of the algorithm, an additional decision mechanism was designed for the DAID units. Three ADALINES were used, and 150 cycles were executed. The classification performance of this system was 94.35, and the generalization performance was 89.5. Both figures are noticeably higher than those in Figure 23.

It was suspected that the highest generalization performance did not necessarily coincide with the highest classification performance for the iterative design technique. The program was modified to permit testing

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		Crater	Linear	Tot.	Crater	Linear	Tot.	Crater	Linear	Tot.
Forced L.		77.9	75.2	76.55	78.5	76.5	77.50	74.0	82.5	78.25
Bayes W.		77.9	75.2	76.55	78.5	76.5	77.50	74.0	83.0	78.50
Error Corr.	96	84.7	98.8	91.75	38.0	79.5	58.75	76.5	73.5	75.00
		96.9	97.3	97.10						
Iter. Des.	26	97.1	97.4	97.25	75.5	74.0	74.75	79.0	71.5	75.25
Mean Sq. Er.	23	90.7	91.7	91.20	75.0	70.5	73.00	68.5	78.5	73.50
Madaline	30	88.5	86.3	87.40	80.0	76.5	78.25			

C1820

Figure 21. Craters vs. Linear Features (1000 SDA Logic Units)

Technique	Cycles	% Design Pats.			% Gen. Pats.			Best % Gen. Pats.		
		Crater	Linear	Tot.	Crater	Linear	Tot.	Crater	Linear	Tot.
Forced L.		94.9	62.1	78.50	55.0	94.5	74.75	92.0	66.5	79.25
Bayes W.		96.3	68.9	82.60	64.0	91.0	77.50	88.0	80.0	84.00
Error Corr.	395	100	89.6	94.80	100	49.5	74.75	98.5	100	99.25
		99.9	98.5	99.20						
Iter. Des.	60	99.7	99.3	99.50	99.5	99.5	99.50	99.5	100	99.75
Mean Sq. Er.	60	98.1	68.4	83.25	68.5	96.0	82.25	79.5	91.5	85.50
Madaline	50	96.7	78.0	87.35	77.0	91.5	84.25			

Figure 22. Craters vs. Linear Features (310 DAID Logic Units)

Technique	% Design Pats.		% Gen. Pats.		Best % Gen. Pats.	
	SDA	DAID	SDA	DAID	SDA	DAID
Forced L.	76.55	78.50	77.50	74.75	78.25	79.25
Bayes W.	76.55	82.60	77.50	77.50	78.50	84.00
Error Corr.	91.75	94.80	58.75	74.75	75.00	99.25
	97.10	99.20				
Iter. Des.	97.25	99.50	74.75	99.50	75.25	99.75
Mean Sq.Er.	91.20	83.25	73.00	82.25	73.50	85.50
Madaline	87.40	87.35	78.25	84.25		

c/822

Figure 23. Craters vs. Linear Features

the generalization performance at the end of each cycle. The program was then run using the SDA units. The resulting classification and generalization error rates are presented on a cycle by cycle basis in Figure 24. It can be seen that if the design was terminated at the end of 7 cycles (instead of 26), a four percent decrease (from 26 to 22) in the generalization error rate is possible, at the expense of a 4.6 percent increase in the classification error rate (from 2.9 to 7.5). All four of the recursive programs will be modified to permit this presentation for the experiments on NIMBUS data.

4.4 Conclusions

This section summarizes the conclusions drawn from the results of the design of recognition systems for the classification of lunar terrain features.

The quadratic property filters (DAID units) provide significantly higher performance levels than do the linear property filters (SDA) units), although the total number of adjustable parameters in the specification of the units was about the same for both property filter sets. Three factors may contribute to this result: the quadratic units are based on contrasts in the patterns as well as average brightnesses (the basis of the linear units); selection was used with the DAID units, that is, only the best 10 percent of the units generated were used; and the DAID set was constructed sequentially, with later units selected to complement the earlier ones.

When the generalization performance is low, Bayes weights and forced learning give the best results with the DAID units. At intermediate performance levels, all of the techniques are approximately equal. At performance levels high enough for an operational system, iterative design is the top performer. Allowing for the deleterious effects of the threshold substitution, error correction seems to be a match for iterative design in this case.

The disappointing performance of the MADALINE appears to be due to the inefficiency of the algorithm coupled with the limitation on the design time. The algorithm is an extension of the error correction method, which is also inefficient. The inefficiency is not as serious for the error correction technique, since the more limited structure permits many more cycles to be accomplished in the allotted design time. It is also possible, in common with

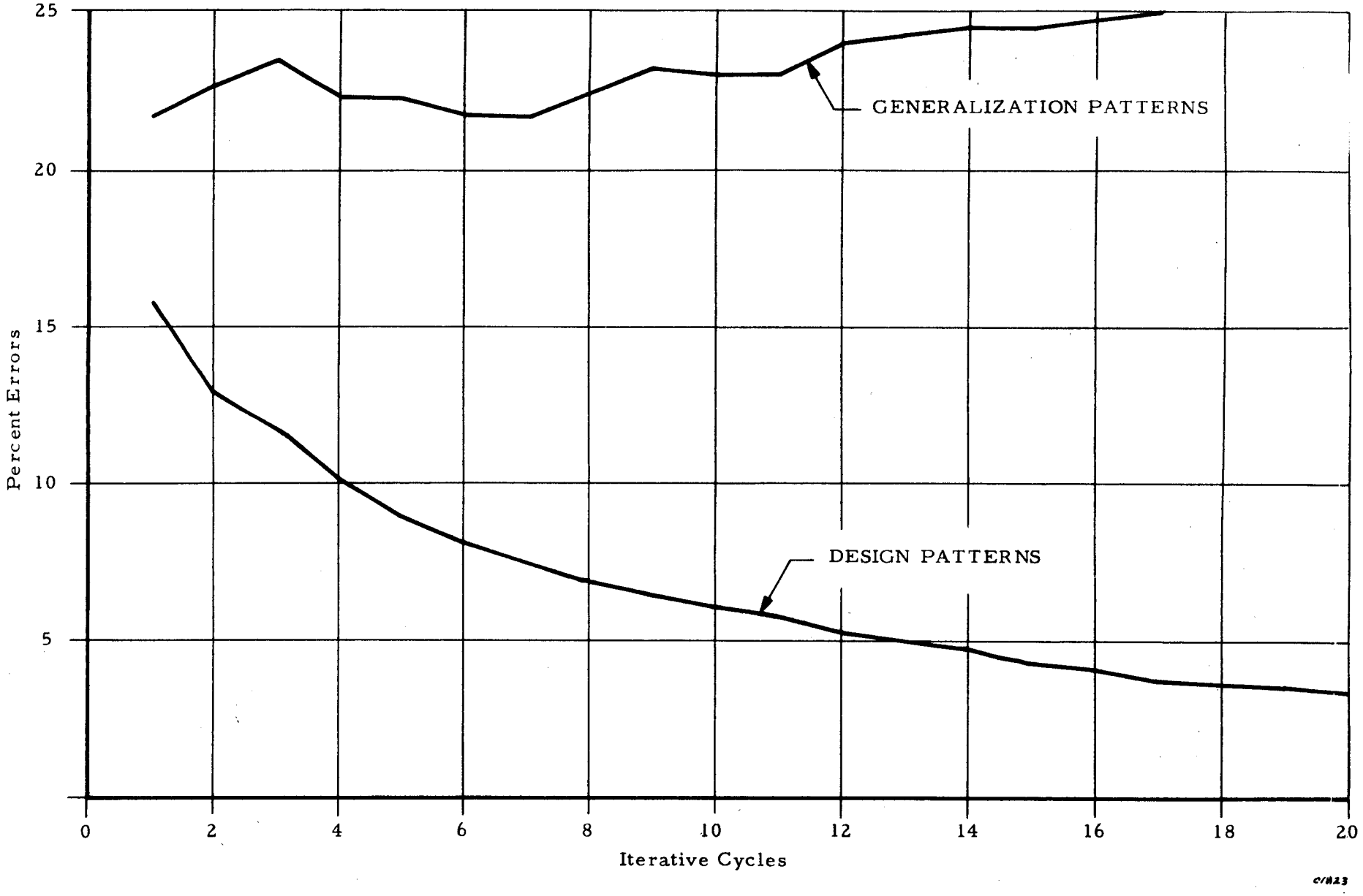


Figure 24. Craters vs. Linear Features, Iterative Design (1000 SDA Units)

error correction, that performance with MADALINE oscillates strongly as a function of the number of cycles, and that unfortunate stopping points were chosen. If this is the case, it will be evident in the NIMBUS experiments, as performance will be computed at the end of each cycle.

The performance obtained appears to be very strongly related to the size of the significant pattern features relative to the size of the sensory field. When the significant feature represents less than one percent of the field, the best generalization performance achieved was 64 percent. When the feature was five to ten percent of the field, the best generalization was 76 percent. With features covering 30 percent of the field, the generalization performance better than 99 percent was achieved.

5.0 PROGRAM FOR THE NEXT PERIOD

During the final reporting period, work will be completed on the experimental program and on the hardware feasibility study. To complete the study of the lunar features, the generalization programs for error correction will be rerun using the threshold designed by the algorithm, rather than the substitute threshold. The error correction program will be modified to remove the threshold substitution, and the programs for error correction, mean square error, and MADALINE will be modified to compute performance at the end of each cycle. The programs will then be run on two recognition tasks using the NIMBUS data--separating the cumulus from noncumulus cloud cover, and separating cumulus solid cells from cumulus polygonal cells. These experiments will be run with the quadratic DAID units only.

One or more hardware implementations of a recognition system will be studied for feasibility. Wherever possible off the shelf hardware components will be used. The functions to be implemented include the optical input system, electro-optical dissection of the image, and sampling of the dissected image, as well as an implementation of the recognition device itself. An estimate for the system weight, size, and power requirements will be attempted. Finally, the transmission bit rate, power, and bandwidth required for transmission of the classification of the dissected image will be compared with that of transmitting the entire image.

REFERENCES

1. Research on the Utilization of Pattern Recognition Techniques to Identify and Classify Objects in Video Data, Astropower Laboratory Report SM-48464-TPR-1, September 1965.

2. Lunar Atlas, Vol. 1, First Edition, University of Chicago, February 1960.

3. Whitaker, E. A., G. P. Kuiper, W. K. Hartmann, and L. H. Spradley, 1963 Rectified Lunar Atlas, Supplement Number 2 to the USAF Lunar Atlas, University of Arizona.

4. Rosenblatt, F., Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms, Spartan Books, Washington, D. C., 1962.

5. Widrow, B., "Generalization and Information Storage in Networks of Adaline Neurons," Self Organizing Systems 1962, Spartan Books, 1962.

6. Highleyman, W. H., "Linear Decision Functions, with Application to Pattern Recognition," Proc. IRE, 50:6, June 1962.

7. Daly, J. A., R. D. Joseph, and D. M. Ramsey, "An Iterative Design Technique for Pattern Classification Logic," 1963 Proc. WESCON Convention, Paper 1.3.

8. Mosteller, F. and D. L. Wallace, Inference & Disputed Authorship: The Federalist, Addison-Wesley Publishing Company, Inc., 1964.

APPENDIX A

SPECIFICATIONS FOR SLOW SCAN TELEVISION SYSTEM MODEL 6030B

I. SLOW SCAN CAMERA.

A. Definition Generators.

1. Line.

Frequency (two ranges)	10 to 1,000 cps
Input sweep sensitivity	2 volts P-P
Sweep Voltage Output	2 volts P-P
Input sync	4 volts negative
Output sync	4 volts negative

2. Frame.

Frequency (two ranges)	.02 to 2 cps
Input sweep sensitivity	2 volts P-P
Sweep voltage output	2 volts P-P
Input sync	4 volts negative
Output sync	4 volts negative

B. Blanking Generator Output, +10 volts (blanking negative).

C. Video.

1.	Frequency Response	1 cps to 1 MC
2.	Output Amplitude	1.0 volt nominal

D.	Resolution	600 lines
E.	Number of Gray Scales	Eight
F.	Sensitivity	0.5 foot-candles faceplate illumination

II. SLOW SCAN MONITOR.

A. Sweep Amplifiers.

Input Sweep Sensitivity	2 volts P-P

B. Video Amplifier.

Frequency response	1 cps to 1 MC
Input sensitivity	1. 0 volt

C. Shutter Control.

Maximum current	10 amp DC

D. Cathode Ray Tube. 5 CK

Spot Size	. 001 IN. at 25uA
Usable Diameter	4. 25 inches
Phosphor	P 12

III. POWER.

A. Input voltage	105-125 VAC, 60 cps
B. Power	150 watts

IV. SIZE AND WEIGHT.

Rack mounted enclosed cabinet.

Height	35-1/4 inches
Width	19-1/2 inches
Depth	20-1/2 inches
Weight, complete	104 pounds